HW5_STAT5000

Xingyu Chen

October 04, 2021

Due through Canvas on Wednesday October 6, 2021. There are 16 lettered parts,sub-lettered parts, or stand-alone questions, so your self-grade total should be 30.

For clarification, these are: Theory:1a,1b,2,3,4a(i),4a(ii),4b,5 + Comp:1,2a,2b,2c,2d,2e,3a,3b. Note: #4a has two sub-lettered parts, namely 4a(i) and 4a(ii), please grade each of these are 2 point questions.

Total Score: 28/30

1 Theoretical Questions

1. Suppose that the number of plants of a particular type found in a rectangular sampling region (called a quadrat by ecologists) in a certain geographic area is a random variable X with pmf

$$P(X = x) = \begin{cases} \frac{c}{x^a} x = 1, 2, 3 \\ 0 otherwises. \end{cases}$$

(a) Let a = 3. Is E(X) finite? Justify your answer. Q1: 2/2

Answer:
$E(x) = \sum_{x=1}^{\infty} p(x) * x$
$= \sum_{x=1}^{\infty} \frac{c}{x^2}$
According to the comparison test, $\sum_{x=1}^{\infty} \frac{c}{x^2}$ converges.
Hence E(X) is finite when a = 3.

(b) Let a = 2. Is E(X) finite? Justify your answer. Q2: 2/2

Answer:
$E(x) = \sum_{x=1}^{\infty} p(x) * x$
$= \sum_{x=1}^{\infty} \frac{c}{x}$
According to the comparison test, $\sum_{x=1}^{\infty} \frac{c}{x}$ diverges.
Hence E(X) is finite when a = 2.

2. Let X = the outcome when a fair die is rolled once. Suppose that, before the die is rolled, you are offered a choice: Option #1: a guarantee of 1/3.5 dollars (whatever the

outcome of the roll); Option #2: h(X) = 1/X dollars. Which option would you prefer? Justify your answer. Q3: 2/2

Answer:
For Option #1:
E(x) = 1/3.5 = 0.2857
For Option #2:
E(x) = 1/6 (1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6) = 0.4083
Thus, Option #2 is better.

3. Recall the distribution from HW #4: X = the leading digit of a randomly selected number from a large accounting ledger. The pmf was defined by:

$$P(X = x) = f(x) = \log_{10}(\frac{x+1}{x})x = 1, 2, ..., 9.)$$

Find E(X). Q4: 2/2

Answer:
$E(X) = log(2/1)*1+log(3/2)*2+log(4/3)*3+log(5/4)*4+log(6/5)*5+log(7/6)*6+log(8/7)*7+log(9/8)*8+log(10/9)*9 = 3.44$

4. (a) Suppose that E(X) = 5 and E[X(X - 1)] = 27.5. What is:

i. $E(X^2)$? Q5: 2/2

Answer:
$E[X(X-1)] = E[X^2 - X] = E[X^2] - E[X] = E[X^2] - 5 = 27.5$ $E[X^2] = 32.5$

ii. Var(X)? Q6: 2/2

Answer:
Var(X) = $E[X^2] - E^2[X]$
$= 32.5 - 5^2$
$= 7.5$

(b) Suppose that a <= X <= b. Prove that a <= E(X) <= b (consider both the discrete and the continuous cases). Q7: 2/2

Answer:
For discrete case: $a <= x <= b$
$a * f(x) <= xf(x) <= b * f(x)$
$\sum a * f(x) <= \sum xf(x) <= \sum b * f(x)$
$a \sum f(x) <= \sum xf(x) <= b \sum f(x)$
$a <= \sum x * f(x) <= b$
$a <= E[X] <= b$

For continuous case:
$E[X] = \int_a^b x * \frac{1}{b-a} = \frac{a+b}{2}$
Thus, a <= E(X) <= b

Thus, proved.

5. Using the definition E(X) = $\sum_{x=0}^{\infty} xP(X = x)$, show that if X $P(\lambda)$, then $E(X) = \lambda$ (HINT: What is the series representation of e^{λ}? HINT #2: I did this in class!). Q8: 2/2

Answer:
Taylor Series: $e^x = \sum_{i=0}^{\infty} \frac{\lambda^c}{x!} = 1 + \lambda + \frac{\lambda^2}{2} + ...$
$E[X] = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} = \lambda$

2 Computational Questions

Instructions for "computational" questions: Your work should be neatly done and in- clude all graphs, code, and comments, labeled and in order based on the problem you are addressing. Do not put graphs in at the end, stick code in random locations, or do anything else that will make this homework difficult to read and grade. If you turn in something that is messy or out of order, it will be returned to you with a zero.

1. Perform a simulation to further justify your answer to theoretical question 2. The simulation should answer the question: which option produces the better outcome if you were offered the choice over and over again? Q9: 2/2
Answer:

```
1/3.5
```

```
## [1] 0.2857143
```

```
mean(1/sample.int(6,10000, replace = TRUE))
```

```
## [1] 0.4079133
```

Thus, Option #2 is better.

2. One really cool (and useful!) application of random variables is approximating integrals/the area under a curve. Consider f(x) = e^x on the interval 0 <= x <= 1. Let's use uniform random variables to approximate the area under f(x) on this interval. Note that this general idea is used often to solve really important but hard integrals!

(a) By hand, and using the integrate() function in R, calculate the true area under f(x). Q10: 2/2
Answer:

```
integrate(function(x){exp(x)}, 0 , 1)
```

```
## 1.718282 with absolute error < 1.9e-14
```

(b) Generate n = 5,000 uniform random (x,y) coordinates in the rectangle x ∈ [0,1], y ∈ [0,e]. Q11: 2/2
Answer:

```
x = runif(5000, min = 0, max = 1)
y = runif(5000, min = 0, max = exp(1))
sample <- cbind(x, y)
```

(c) Calculate the proportion of points from (b) that fall below f(x) and use this proportion to approximate the area under f(x). Q12: 2/2

Answer:

```
library(dplyr)
sample <- as.data.frame(sample)
sample_below <- filter(sample, y < exp(x))
proportion <- as.numeric(count(sample_below)/5000)
proportion
```

```
## [1] 0.6388
```

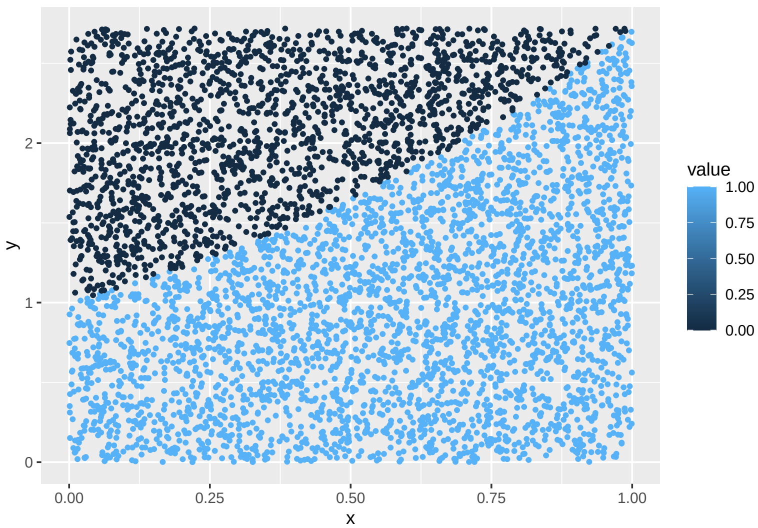
```
exp(1) * proportion
```

```
## [1] 1.736438
```

(d) Plot all of the randomly generated points, giving different colors for point above and below f(x). Q13: 2/2

Answer:

```
library(tidyverse)
sample_below$value = 1
sample_up <- filter(sample, y > exp(x))
sample_up$value = 0
sample_graph <- rbind(sample_below,sample_up)
ggplot(data = sample_graph, aes(x=x, y=y, color = value)) +
  geom_point(size=1)
```



(e) Find the error between our approximation and the true area calculated in (a). How can we make this error smaller? Q14: 2/2

Answer:

```
abs(as.numeric(exp(1) * proportion) - 1.718282)
```

```
## [1] 0.01815643
```

We can use more points instead of just 5000 points.

3. This question pertains to the [standard] Beta distribution:

(a) In the same window, plot four different beta distributions (pdf):

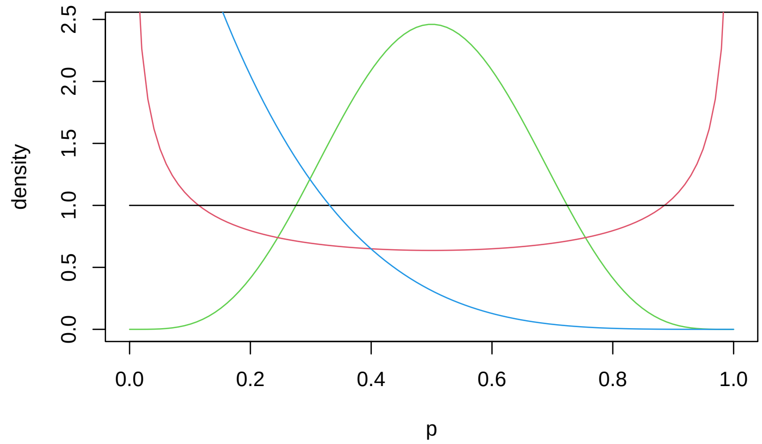
i. beta(1,1)

ii. beta(0.5,0.5)

iii. beta(5,5)

iv. beta(1,5)
What is another name for (i)? Q15: 1/2
Answer:

```
p = seq(0,1, length=100)
plot(p, dbeta(p, 5, 5), ylab="density", type ="l", col=3)
lines(p, dbeta(p, 0.5, 0.5), col=2)
lines(p, dbeta(p, 1, 1), col=1)
lines(p, dbeta(p, 1, 5), type ="l", col=4)
legend(0.7,8, c("Be(1,1)","Be(0.5,0.5)","Be(5,5)", "Be(1,5)"),lty=c(1,1,1,1),col=c(4,3,2
```



Other name: the uniform distribution

(b) Now, suppose your friend has a coin whose probabilities are unknown to you. Since the standard beta distribution has support [0,1], it is well-suited to model your beliefs about the probability of heads. Match the following descriptions with the beta distributions from the previous part (Computational #3(a)).

i. I am pretty confident that this is a two-headed or two-tailed coin.

ii. I have no idea what the probability of heads is.

iii. I believe that the probability of heads is low.

iv. I believe that the coin is close to fair. Q16: 1/2

Answer:

(b) i <- (a) ii
(c) ii <- (a) iii
(d) ii <- (a) iv
(e) iv <- (a) i